AD_____

GRANT NUMBER DAMD17-94-J-4386

TITLE: Detection and Characterization of Autoantigens in Breast Cancer

PRINCIPAL INVESTIGATOR: Janis Racevskis, Ph.D.

CONTRACTING ORGANIZATION: Montefiore Medical Center
Bronx, New York 10467

REPORT DATE: August 1997

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

RACEVSKIS, Janis

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1997	Annual (15 Jul 96 - 14 Jul 97)

4. TITLE AND SUBTITLE
Detection and Characterization of Autoantigens in Breast Cancer

5. FUNDING NUMBERS
DAMD17-94-J-4386

6. AUTHOR(S)
Janis Racevskis, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Montefiore Medical Center
Bronx, New York 10467

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, MD 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

The entire sequence of our second breast tumor associated autoantigen (*Auag2*), which is a newly discovered gene product, has been determined. The predicted amino acid sequence of *Auag2* contains two LIM domain motifs, which are conserved cysteine rich zinc-binding motifs of about 60 amino acid residues that mediate protein-protein interactions, and are characteristic of a group of critical transcriptional regulators of embryonic development. The 350 bp region coding for the two tandem LIM domains was found to be 60% homologous to the analogous region of rhombotin 1, a proto-oncogene of 160 amino acids whose gene is disrupted by chromosomal translocation in T-cell leukemia. A third autoantigen clone has been identified and is being characterized. Various features of the *Auag2* cDNA sequence (a long GC-rich structured 5′ end, the presence of mRNA destabilizing motifs in the 3′ end and a homology to a known oncogene) all predict that this gene plays a vital role in the life of the organism, and may play a role in malignancy.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	14
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

JR Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

JR For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

JR In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

JR In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Jarin Raczynski 9/9/97

PI - Signature Date

TABLE OF CONTENTS

INTRODUCTION

Tumor growth is associated with the expression of mutated gene products, inappropriate gene expression, and the breakdown of tissue architecture, leading to the exposure and release into the peripheral circulation of sequestered antigens (1,2). Whether these circulating, mutated or newly displayed tumor-associated antigens elicit an autologous humoral immune response in the breast tumor patient is of vital interest. Isolation, identification and characterization of novel breast tumor associated autoantigens might yield new insights into the disease process, and moreover, may be developed into diagnostic screening tests and potential targets for immunotherapy.

The screening of cDNA expression libraries with autologous patient serum is a powerful technique, which has been used successfully for the identification of autoimmune disease antigens (3), and which we have adapted for the identification of autoantigens in cDNA libraries made from breast tumor mRNA. After screening cDNA libraries, derived from primary ductal breast carcinomas with autologous patient serum, we have detected and isolated three immunoreactive cDNA clones, all three of which are newly discovered gene products. The first autoantigen isolate *Ngp 1* has been characterized and is a nucleolar GTP-binding protein. The predicted amino acid sequence of the second clone *Auag2* contains two LIM domain motifs and bears a 60% homology in this region to a known oncogene, Rhombotin 1. Sequencing of the third isolate *Auag3* is underway. Our studies have identified novel proteins that might be involved in malignancy, and may help understand the biology of breast cancer.

BODY

Most of our effort for the past year has focused on characterizing our second breast tumor autoantigen isolate (working name *Auag2*), which is also a newly discovered gene. We chose to focus on *Auag2* because we found it to be related to a group of known oncogenes, and it may represent an important new discovery. Work is continuing on our first autoantigen isolate *Ngp-1*, the nucleolar GTP-binding protein (4), to identify other proteins that interact with it using the yeast two-hybrid vector system (5). We have encountered difficulties using the original phagemid vector *pBD-GAL4* as the bait plasmid expressing the open reading frame of *Ngp-1*, and will continue this work using recently available more tightly controlled inducible vectors (*pGILDA*). The inducible vectors allow performing two-hybrid library screening with bait proteins that interact with yeast host cell proteins, or are toxic to yeast cells. The highly conserved nature of *Ngp-1* which we even found to have significant homology to a gene from such a distant organism as rice (4), probably accounts for the difficulties we encountered. We have isolated a third autoantigenic clone fragment (*Auag3*) which, with the exception of expressed sequence tags in the databases, shows no homology to any known gene (Figure 5, page 14). A 3kb clone containing *Auag3* sequences has been isolated and is being characterized.

Auag2 cDNA contains a long region of extremely high GC content in the 5′ third of the molecule (Figure 1, page 11) which interferes with reverse transcription, hence full length clones of *Auag2* are under-represented in cDNA libraries. This was illustrated by our cDNA homology searches of the databases, which identified a number of identical human expressed sequence tags, however, none of them representing the first 900 bases of the *Auag2* cDNA. Furthermore, we found that standard PCR reactions could not amplify the GC rich region, and special formulations which melt high GC DNA had to be employed. A full length clone containing 2.1 kb of *Auag2* sequence was isolated by using the Gene Trapper technology, where a cDNA library in a plasmid vector is converted to single strand form, then hybridized with gene specific biotinylated oligonucleotide probes which are captured by avidin coated magnetic beads, thus highly enriching for the desired gene product. The high GC content also made accurate sequencing a difficult task and required the use of dITP in the sequencing reactions and formamide sequencing gels. Clones of *Auag2* have been isolated from two different cDNA libraries (brain and testis), and the entire sequence determined accurately (*Genbank accession # U24576*). The first ATG codon in the preferred configuration with an A in position -3, and a G in position +4 (6), is located at position 781 at the start of an open reading frame which could encode a protein of 165 amino acids (Figure 1, page 11). The predicted amino acid sequence of *Auag2* contains two LIM domain motifs (Figure 2, page 12), which are conserved cysteine rich zinc-binding motifs of about 60 amino acid residues that mediate protein-protein interactions, and are

characteristic of a group of critical transcriptional regulators of embryonic development (7). A search of the databases revealed that the 350 bp region coding for the two tandem LIM domains was 60% homologous to the analogous region of rhombotin 1 (Figure 3, page 13), a proto-oncogene of 160 amino acids whose gene is disrupted by chromosomal translocation in T-cell leukemia (8,9). At the amino acid level, the homology, and spacing of the amino acids making up the LIM domains is even more apparent (Figure 4, page 13). These sequences are so highly conserved that the homology to the rhombotin 1 homologue of Drosophila is just as extensive (Figure 4, page 13). A number of the expressed sequence tags found in homology searches were from mouse cDNA libraries and showed a greater than 97% homology, also indicating that *Auag2* is highly conserved in evolution. Although the long untranslated 5′ end of *Auag2* shows a slight homology to some other genes with GC rich regions such as vascular endothelial growth factor, the sequence is unique, as is the long 3′ end. *Auag2* does not seem to have any other closely related genes, since after extensive screening of different cDNA libraries with cDNA probes, in our efforts to obtain full length clones, we have not isolated any related cDNAs. We did however isolate one 5′ end splice variant numerous times. The most extensively studied of the rhombotins (three known members) is rhombotin 2, which has also been shown to be a proto-oncogene in T-cells, and act as a transcriptional regulator of erythroid development (9,10). Rhombotin 2 is widely distributed in many tissues with the exception of T-lymphocytes, and is believed to have multiple, as yet unknown regulatory functions. As judged by northern blot analysis, the distribution of *Auag2* expression is unlike that of RBTN1 or RBTN2. In addition, the long, unique, 5′ and 3′ regions of *Auag2* suggest that *Auag2* has different control elements specifying other tissue distribution and regulatory functions.

The extremely long 5′ end of *Auag2* (780 bp) is the GC rich region (Figure 1, page 11). A long GC rich, structured 5′-leader sequence is characteristic of transcripts encoding oncoproteins, growth factors, transcription factors, and other regulatory proteins - that seem to be designed to be translated poorly (11). Inhibition at the translational level seems to be a component of gene regulation for genes which need to be tightly regulated. Another feature of the sequence of *Auag2* cDNA is the presence of multiple ATTT motifs in the 3′ end (Figure 1), which have also been observed in the 3′ untranslated region of numerous lymphokine, cytokine, and proto-oncogene mRNAs. It has been proposed that such ATTT motifs are involved in the selective degradation of transiently expressed messengers (12). In northern blots of mRNAs isolated from various human tissues, *Auag2* appears to be most highly expressed in testes and brain, and is not detectable in liver and kidney. *Auag2* transcripts are also present in mRNA extracted from breast tumors, however, since breast tumors are a complex mixture of different cell types (stromal fibroblasts, infiltrating lymphocytes and transformed breast epithelial cells), we will have to do immunohistochemistry and in-situ hybridization to determine the

exact source of these transcripts. In certain tissues there appears to be an extra band; probably representing different splicing products. A splice variant involving the 5' end of *Auag2* cDNA has been isolated from different cDNA libraries; its role in the expression of *Auag2* protein is yet to be determined. Because of the highly conserved nature of the *Auag2* protein, it may prove to be a weak immunogen in rabbits. An alternative is to use the anti-peptide antibody approach which enables the animal host to bypass the immune tolerance frequently encountered when a highly conserved or self protein antigen is used as the immunogen. We therefore have obtained antiserum against a synthetic peptide from the open reading frame amino terminus region, outside the LIM domain, which has no homology to the rhombotins, for immunohistochemical localization of *Auag2* protein within different tissues and cell types.

Since LIM domain containing proteins interact with other proteins to form specific transcription regulators, we will try to identify those that interact with *Auag2* by applying the yeast two hybrid system. Rhombotin 2, for example, has been shown to interact with retinoblastoma-binding protein 2 (10).

CONCLUSIONS

Clones of *Auag2* have been isolated from two different cDNA libraries and the entire sequence determined accurately. The predicted amino acid sequence of *Auag2* contains two LIM domain motifs, which are conserved cysteine rich zinc-binding motifs of about 60 amino acid residues that mediate protein-protein interactions, and are characteristic of a group of critical transcriptional regulators of embryonic development. A search of the databases revealed that the 350 bp region coding for the two tandem LIM domains was 60% homologous to the analogous region of rhombotin 1, a proto-oncogene of 160 amino acids whose gene is disrupted by chromosomal translocation in T-cell leukemia. As judged by northern blot analysis, the distribution of *Auag2* expression is unlike that of RBTN1 or RBTN2. In addition, the long, unique, 5′ and 3′ regions of *Auag2* suggest that *Auag2* has different control elements specifying other tissue distribution and regulatory functions.

Various features of the *Auag2* cDNA sequence (a long GC-rich structured 5′ end, the presence of mRNA destabilizing motifs in the 3′ end and a predicted amino acid sequence which contains two LIM domain motifs with a partial homology to a known oncogene) all predict that this gene plays a vital role in the life of the organism, and merits further investigation and characterization.

REFERENCES

1. Naftzger, C., and Houghton, A.N. Tumor immunology. Current Opinion in Oncology, *3*:93-99, 1991.

2. Henderson, R.A., and Finn, O.J. Human tumor antigens are ready to fly. Advances in Immunology, 62:217-251, 1996.

3. Tan, E.M. Autoantibodies in pathology and cell biology. Cell, *67*:841-842, 1991.

4. Racevskis, J., Dill, A., Stockert, R., and Fineberg, S.A. Cloning of a novel nucleolar guanosine 5'-triphosphate binding protein autoantigen from a breast tumor. Cell Growth & Differentiation, 7:271-280, 1996.

5. Fields, S., and Ok-kyu Song. A novel genetic system to detect protein-protein interactions. Nature (London), *340*:245-247, 1989.

6. Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. *15*:8125-8148, 1987.

7. Sanchez-Garcia, I., and Rabbits, T.H. The LIM domain: a new structural motif found in zinc-finger-like proteins. Trends Genet. *10*:315-320, 1994.

8. NcGuire, E.A., Hockett, R.D., Pollock, K.M., Bartholdi, M.F., O'Brien, S.J., and Korsmeyer, S.J. The t(11;14)(p15;q11) in a T-cell acute lymphocytic leukemia cell line activates multiple transcripts, including Ttg-1, a gene encoding a potential zinc finger protein. Mol. Cell. Biol. *9*:2124-2132, 1989.

9. Boehm, T., Foroni, L., Kaneko, Y., Perutz, M.F., and Rabbits, T.H. The rhombotin family of cysteine-rich Lim-domain oncogenes: distinct members are involved in T-cell translocations to human chromosomes 11p15 and 11p13. Proc Natl. Acad. Sci. USA, *88*:4367-4371, 1991.

10. Mao, S., Neale, G.A.M., and Goorha, R.M. T-cell oncogene rhombotin-2 interacts with retinoblastoma-binding protein 2. Oncogene, *14*:1531-1539, 1997.

11. Kozak, M. An analysis of vertebrate mRNA sequences: intimations of translational control. J. Cell Biol., *115*:887-903, 1991.

12. Shaw, G., and Kamen, R. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell, *46*:659-667, 1986.

```
1    GCTCTGTCAGTAACACATGTGTAAGAGCCGCGGAGGGAGCGAGCGAGCCGGCTAGAGGCC
61   AGCGCCGCCGCCGCCGCCGCCTCCGAGCCGGGCAGCAACAGCCCCGGCAGCGGCGCAGGC
121  TCCAGCGCGCCGGGCCCGGCCGGCCGCAGCCCCCGACGCCTGGGTGCGCCTGCCTGCCGG
181  CCTCCGCACCGTCCGCCGCCGCTCCCGGGGCTGTTGTGTCTGCGACTGCTCCCGGCCGGA
241  GGTGCAGGGAGCTCAGCCGAGCCGCCGCTGCCATCCCGGAGCGAGCAAGCGAGCGAGCGC
301  GCGGGAGGGAGGAAGGCGGCGGCGGAGGAGGAGGAGGAGCGGGAGGAGCGCGGGCGGGGG
361  CGGGGGCCGCCGGGCGGGGGGAATATACAAAGTGAAGCCACATTGCCAAACTTGCAGCAGC
421  GATTGCAGCAGTTGCTGCCGCTGCGCCGCGCCTGAAGCCGCGCCGCGCGGGCCGAGGGCT
481  CCTGCAGCTGCTCGCGCGCAGTCGGAGGCGGAGAAGGACGAAGACTGAGACTGACACTTC
541  TGCTCCCGGCCGCCCGGCACTTACGCGGGGGCCCCCCAACCCGCCCCAGAGCAACGCGAT
601  TTAAAAAAAAAAAAAAAAGCCGCCCTTAGCCCCCTCCTCCCCTTTCCTGCTTCTGCGAGAA
661  CTCCCTCCCTCCCTCCAGCTCCGCCAGCCCAGGCGCCCCTTCCCTGGAAGCCGAGCGGCT
721  TCGCTCGCATTTCACCGCCGCCGCCTCTCGCAATATTGCAATATAGGGGAAAAGCAGACC
781  ATGGTGAATCCGGGCAGCAGCTCGCAGCCGCCCCCGGTGACGGCCGGCTCCCTCTCCTGG
841  AAGCGGTGCGCAGGCTGCGGGGGCAAGATTGCGGACCGCTTTCTGCTCTATGCCATGGAC
901  AGCTATTGGCACAGCCGGTGCCTCAAGTGCTCCTGCTGCCAGGCGCAGCTGGGCGACATC
961  GGCACGTCCTGTTACACCAAAAGTGGCATGATCCTTTGCAGAAATGACTACATTAGGTTA
1021 TTTGGAAATAGCGGTGCTTGCAGCGCTTGCGGACAGTCGATTCCTGCGAGTGAACTCGTC
1081 ATGAGGGCGCAAGGCAATGTGTATCATCTTAAGTGTTTTACATGCTCTACCTGCCGGAAT
1141 CGCCTGGTCCCGGGAGATCGGTTTCACTACATCAATGGCAGTTTATTTTGTGAACATGAT
1201 AGACCTACAGCTCTCATCAATGGCCATTTGAATTCACTTCAGAGCAATCCACTACTGCCA
1261 GACCAGAAGGTCTGCTAAAGGTCAGAGTAATGCAGAATGCGTGCCTTCATCTCAGATTT
1321 GTTCATCACAGGTGGATCCCATGTGTCTTCAGTAGACAAGTCACCTTTGTAGCTAGCACC
1381 AGTGCCAGCTCCATGCCATTGCACCTTCTTTAGTCTTGATTGCCCTTCCCGCATTTATTG
1441 GTGTATTAAAATGACTGAATATGAACATTAAGGACTCCATGAACCTGGGCTAATGGGAGA
1501 CTGTAGAGAAAATGAAAAAAGATCCACCAGAGGACATCTTGGGGAGGGGGAGGGAGCTGG
1561 GGGGGAGGGAAATGACTAATGAAGCTAATTAAAAGAAGCATTCAAATCTGCTTTCTACCC
1621 TCATTAACAATTAGCAGGGCACTGGCCAGAGTTTGTACCCTGTGTTTTACCTTAACAACA
1681 TTCTATTTGCTCTTTGTATATTTAAGTGTTGTAAGGAAACGTGTTTCAATCAAAACTGAC
1741 CATGAGATAAAGGAAAGAGATGTGGCTTTTGTGATATTCTATCACAAACACTTATTGTAT
1801 CTCTGTAAAATACAATGTATGTATGCATGTAAGTGTTTTTGTCCTAATGTTGCTACTCCC
1861 ATGGCAAAGAAAAAAAAAAGAATGAAAAAAAGAAAAAAAATTTGGAAAAAAAAATCAGGC
1921 TCATAGCAGCTACTGTGTAGAAAATTCCCCCTACTTCTAATTTGCTGAATGAAGAAAAAA
1981 AAAAATCTTTTATTTGTGATATTTTCAGAGACATTTGCTCTAGTATGGTGTATTTAAATA
2041 ATAAAAACTTAAAAGAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
2101 AAAAAAAAAAAAAAAAAAAAAAAAA 3'
```

Figure 1. Complete nucleotide sequence of *Auag2* cDNA (Genbank accession # U24576). Initiation (nucleotide **781**) and termination codons (nucleotide **1278**) of the putative open reading frame are shown in bold double underline. G and C residues in the 5' untranslated region (nucleotides 1-780) are shaded, and ATTT motifs in the 3' region are underlined. Potential polyadenilation signal **AATAAA** is in bold at position 2040.

```
   1   M  V  N  P  G  S  S  S  Q  P  P  P  V  T  A  G  S  L  S  W  K
 781   ATGGTGAATCCGGGCAGCAGCTCGCAGCCGCCCCCGGTGACGGCCGGCTCCCTCTCCTGGAAG

  22   R  C  A  G  C  G  G  K  I  A  D  R  F  L  L  Y  A  M  D  S  Y
 844   CGGTGCGCAGGCTGCGGGGGCAAGATTGCGGACCGCTTTCTGCTCTATGCCATGGACAGCTAT

  43   W  H  S  R  C  L  K  C  S  C  C  Q  A  Q  L  G  D  I  G  T  S
 907   TGGCACAGCCGGTGCCTCAAGTGCTCCTGCTGCCAGGCGCAGCTGGGCGACATCGGCACGTCC

  64   C  Y  T  K  S  G  M  I  L  C  R  N  D  Y  I  R  L  F  G  N  S
 970   TGTTACACCAAAAGTGGCATGATCCTTTGCAGAAATGACTACATTAGGTTATTTGGAAATAGC

  85   G  A  C  S  A  C  G  Q  S  I  P  A  S  E  L  V  M  R  A  Q  G
1033   GGTGCTTGCAGCGCTTGCGGACAGTCGATTCCTGCGAGTGAACTCGTCATGAGGGCGCAAGGC

 106   N  V  Y  H  L  K  C  F  T  C  S  T  C  R  N  R  L  V  P  G  D
1096   AATGTGTATCATCTTAAGTGTTTTACATGCTCTACCTGCCGGAATCGCCTGGTCCCGGGAGAT

 127   R  F  H  Y  I  N  G  S  L  F  C  E  H  D  R  P  T  A  L  I  N
1159   CGGTTTCACTACATCAATGGCAGTTTATTTTGTGAACATGATAGACCTACAGCTCTCATCAAT

 148   G  H  L  N  S  L  Q  S  N  P  L  L  P  D  Q  K  V  C  *
1222   GGCCATTTGAATTCACTTCAGAGCAATCCACTACTGCCAGACCAGAAGGTCTGCTAA
```

Figure 2. Predicted amino acid sequence of the *Auag2* open reading frame. LIM domain motif amino acids are shaded and conform to the concensus sequence of all known LIM domains (7):

$$[\mathbb{C}\ X_2\ \mathbb{C}\ \ldots\ X_{16\text{-}23}\ \ldots\ \mathbb{H}\ X_2\ \mathbb{C}\ X_2\ \mathbb{C}\ X_2\ \mathbb{C}\ \ldots\ X_{16\text{-}21}\ \ldots\ \mathbb{C}\ X_{2\text{-}3}\ (\mathbb{C},\mathbb{H},\mathbb{D})]$$

```
                  850        860        870        880        890        900
AUAG2.DNA     AAGCGGTGCGCAGGCTGCGGGGGCAAGATTGCGGACCGCTTTCTGCTCTATGCCATGGAC
              ::: : :: :: X:::: :::::::  :::::::: ::::::  : ::  :::::
RBTN1.DNA     AAGGGCTGTGCGGGCTGTAACCGCAAGATCAAGGACCGCTATCTGCTGAAGGCATTGGAC
                  570        580        590        600        610        620
                  910        920        930        940        950        960
AUAG2.DNA     AGCTATTGGCACAGCCGGTGCCTCAAGTGCTCCTGCTGCCAGGCGCAGCTGGGCGACATC
              :  :: ::::::      ::::::::::  : :::::  :    :  ::::::::  :
RBTN1.DNA     AAGTACTGGCACGAAGACTGCCTCAAGTGTGCGTGCTGTGACTGCCGCCTGGGCGAGGTG
                  630        640        650        660        670        680
                  970        980        990       1000       1010       1020
AUAG2.DNA     GGCACGTCCTGTTACACCAAAAGTGGCATGATCCTTTGCAGAAATGACTACATTAGGTTA
              ::: :  ::   ::::::::      : : ::::: ::: ::   :::::: : ::: :
RBTN1.DNA     GGCTCCACCCTCTACACCAAGGCCAACCTCATCCTGTGCCGACGCGACTACCTGAGGCTC
                  690        700        710        720        730        740
                 1030       1040       1050       1060       1070       1080
AUAG2.DNA     TTTGGAAATAGCGGTGCTTGCAGCGCTTGCGGACAGTCGATTCCTGCGAGTGAACTCGTC
              ::::: :  :  ::    ::    :::::: :  ::  ::: :: ::    ::  : ::
RBTN1.DNA     TTTGGCACCACAGGGAACTGTGCTGCTTGCAGCAAGCTGATCCCAGCCTTCGAGATGGTG
                  750        760        770        780        790        800
                 1090       1100       1110       1120       1130       1140
AUAG2.DNA     ATGAGGGCGCAAGGCAATGTGTATCATCTTAAGTGTTTTACATGCTCTACCTGCCGGAAT
              ::: :::: :  : ::: :::::::: ::  : :: ::  : :::     :::X
RBTN1.DNA     ATGCGGGCCCGGGACAACGTGTATCACCTCGACTGCTTCGCCTGCCAGCTCTGCAACCAG
                  810        820        830        840        850        860
```

Figure 3. Alignment of the region encoding the LIM domains of *Auag2* and Rhombotin 1 (RBTN1).

```
AUAG2   21 KRCAGCGGKIADRFLLYAMDSYWHSRCLKCSCCQAQLGDIGTSCYTKSGMILCRNDYIRL 80
              K CAGC  KI DR+LL A+D YWH  CLKC+CC  +LG++G++ YTK+ +ILCR DY+RL
RBTN1   22 KGCAGCNRKIKDRYLLKALDKYWHEDCLKCACCDCRLGEVGSTLYTKANLILCRRDYLRL 81

AUAG2   81 FGNSGACSACGQSIPASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYINGSLFCEHD 140
              FG +G C+AC + IPA E+VMRA+ NVYHL CF C  C  R   GD+F   N  + C+ D
RBTN1   82 FGTTGNCAACSKLIPAFEMVMRARDNVYHLDCFACQLCNQRFCVGDKFFLKNNMILCQMD 141


AUAG2   23 CAGCGGKIADRFLLYAMDSYWHSRCLKCSCCQAQLGDIGTSCYTKSGMILCRNDYIRLFG 82
              CAGCG  I DR+LL A+D  WH  CLKC CC  +LG++G++ YTK  ++LC+ DY+RLFG
DROS    45 CAGCGKHIQDRYLLRALDMLWHEDCLKCGCCDCRLGEVGSTLYTKGNLMLCKRDYLRLFG 104

AUAG2   83 NSGACSACGQSIPASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYINGSLFCEHDRP 142
              N+G C+AC + IPA E+VMRA+ NVYHL+CF C  C +R   GDRF+     + CE+D
DROS   105 NTGYCAACSKVIPAFEMVMRARTNVYHLECFACQQCNHRFCVGDRFYLCENKILCEYDYE 164
```

Figure 4. Alignment of the amino acid sequences of the LIM domains of *Auag2* with Rhombotin 1 (RBTN1), and Drosophila homologue of RBTN1 (DROS).

```
AUAG3   GGTATCTGGGCAGCCACTGCCAGCATCATCTTTGCTTTCTTGGGATTTGATACCATTGCT 61
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Z16206  GGTATCTGGGCAGCCACTGCCAGCATCATCTTTGCTTTCTTGGGATTTGATACCATTGCT 83

AUAG3   GCACCTGCTGCGGAAGTTAAGAATCCACAGAAAACGATGGCACGTGGTATCATCGGGACG 121
        ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Z16206  GCACATGCTGCGGAAGTTAAGAATCCACAGAAAACGATGGCACGTGGTATCATCGGGACG 143

AUAG3   GTTTTGATTTCAGCCCTACTCTATGTGTTGTTCGCCGTTGTTTTGACTGGTATTGTGAAC 181
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Z16206  GTTTTGATTTCAGCCCTACTCTATGTGTTGTTCGCCGTTGTTTTGACTGGTATTGTGAAC 203

AUAG3   CTATAAAAAGTTGG 195
            ||||  | ||
Z16206  TATAAAAAGTTGGG 217
```

Figure 5. 5′ end sequences of the 3kb isolate of *Auag3* aligned with an expressed sequence tag (Z16206) from a human cDNA library.